. . . .

Web appendix: Practical Marginalized Multilevel Models

Michael E. Griswold[a]*, Bruce J. Swihart[a], Brian S. Caffo[a], Scott L. Zeger[a]

. . . .

A Log-Log-Normal Model

When continuous data are highly skewed, such as patients' monthly medical expenditures, log-linear gamma models are often used to account for the non-normality. A marginal log-linear model for skewed data arising in clusters may be written as:

Log-Log-Normal MMM:

$$
\begin{aligned}
&i) \quad \mu_{ij}^m = \exp(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\
&ii) \quad \mu_{ij}^c = \exp(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a}_i) \\
&iii) \quad \boldsymbol{a}_i \sim \text{MVN}(0, \boldsymbol{D}) \\
&iv) \quad Y_{ij}|\boldsymbol{a}_i \sim \Gamma(\mu_{ij}^c, \upsilon)
\end{aligned}
$$

Where the gamma distribution is parameterized such that $E(Y^c) = \mu^c$ and $\text{var}(Y^c) = (\mu^c)^2/\upsilon$. To estimate $\boldsymbol{\alpha}^m$, we determine Δ_{ij} using the marginalization constraint:

$$
\begin{aligned}
\boldsymbol{\alpha}^m &= (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\log\left\{\int_a \exp(\Delta_{ij} + \boldsymbol{z}_{ij}\boldsymbol{a})\, d\boldsymbol{F}_{\boldsymbol{a}}\right\} \\
&= (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\log\left\{\exp(\Delta_{ij} + \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}/2)\right\}
\end{aligned}
$$

and thus, $\Delta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m - \boldsymbol{z}'_{ij}\boldsymbol{D}\boldsymbol{z}_{ij}/2$.

. . . .

[a] **Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD. 21205, U.S.A.**
*Email: mgriswol@jhsph.edu

. . . .

A Log-Log-Gamma Model

Public health studies frequently involve Poisson processes where counts of incidents in a specified interval are recorded across multiple visits, locations or both. A mixture of the poisson distributions over a gamma process is often used to account for extra variability (overdispersion) observed in count data of this type. The resulting marginal distribution of the mixture is negative-binomial. A log-linear model for such data may be written as:

Log-Log-Gamma MMM:

$$
\begin{aligned}
&i) &&\lambda_{ij}^m = \exp(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) \\
&ii) &&\lambda_{ij}^c = \exp\{\Delta_{ij} + \log(a_i)\} \\
&iii) &&a_i \sim \Gamma(\nu, k) \\
&iv) &&Y_{ij}|a_i \sim \mathrm{Poisson}(\lambda_{ij}^c)
\end{aligned}
$$

Where the gamma parameterization produces $E(a) = \nu$, $\mathrm{var}(a) = \nu^2/k$. Determining Δ_{ij} with the marginalization constraint we have:

$$
\boldsymbol{\alpha}^m = (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\log\left[\int_a \exp\{\Delta_{ij} + \log(a)\}\, d\boldsymbol{F_a}\right] = (\boldsymbol{x}'_{ij}\boldsymbol{x}_{ij})^{-1}\boldsymbol{x}'_{ij}\log\left[\exp\{\Delta_{ij} + \log(\nu)\}\right]
$$

and thus, $\Delta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m - \log(\nu)$. Lee & Nelder (1996) recommend constraining $E(a) = \nu = 1$, implying that the coefficient of variation in the gamma random effects distribution is constant, $\sqrt{\mathrm{var}(a)}/E(a) = 1/\sqrt{k}$. In this case $\Delta_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m$.

The Page Approximation

Solving: $\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m \doteq 2a_1\Delta_{ij}^{page} + 2a_1 a_2(\Delta_{ij}^{page})^3$ yields:

$$
\begin{aligned}
\Delta_{ij}^{page} &= \sqrt[3]{t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) + u(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)} - \sqrt[3]{u(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)} \\
\text{where:} \quad t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) &= \boldsymbol{x}_{ij}\boldsymbol{\alpha}^m/(2a_1 a_2) \\
u(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) &= \tfrac{1}{2}\left\{-t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m) + \sqrt{t(\boldsymbol{x}_{ij}\boldsymbol{\alpha}^m)^2 - 4/(3a_2)^3}\right\}
\end{aligned}
$$

Further Discussion

Computational techniques in existing software have been proposed for MMMs for an audience of a varied skill set. Career statisticians familiar with R and SAS can utilize the procedures provided herein to obtain marginal likelihood-based inference with well known software. Those with less statistical programming expertise can use software of choice to fit a random intercept model and arrive at a marginal estimate through the appropriate rescaling of estimates (as in the PPN model) or implementing the approximations as outlined in Section 3.1. Our MMM specifications provide methods for those with programming backgrounds to utilize existing software for direct estimation, while the approximations allow simple marginalization from conditional output of standard packages for general users.

Conditional model (GLMM) estimates and interpretations can be heavily dependent on assumed variance structures, as shown in Heagerty & Zeger (2000). Marginal models estimate effects that are directly observable in the data and are more robust to the chosen dependence model. This is illustrated by the similarity in the marginal mean parameter estimates across the range of association assumptions in the crossover trial and visual impairment examples of Section 4. While alternative approaches to estimating marginal models, such as a GEE approach, avoid specifying the complete joint distribution of the responses, the MMM approach retains the capability of likelihood inference and

the consequent benefits therein. The GEE approach precludes the benefits of full likelihood-based analysis of data, such as weaker restrictions on missing data processes for valid inferences. For instance, if there were missing data in the visual impairment study which depended on race, the type of missingness would be MAR and would exclude GEE from modeling choice considerations due to its requirement of missing data to be MCAR (Rubin (1976); Robins, Rotnitzky & Zhao (1995); Scharfstein, Rotnitzky & Robins (1999)). Weighted GEE (WGEE) is a well-known fix for GEE in MAR situations, however is difficult to apply with intermittent missingness which poses no additional considerations for the MMM (Molenberghs & Verbeke (2005)).

Simulation studies as well as published data sets stated to have MAR responses were utilized in comparing the estimates of MMM to GEE (results available upon request). No appreciable differences were found in the specific instances tested. In light of the theory pertaining to consistent GEE estimates in MAR situations, the lack of appreciable differences may be due to the method of artificially inducing MAR in the simulation setting or features of the specific data sets used as opposed to vetting GEE in all MAR situations. A suitable simulation comparing GEE and MMM in MAR situations is sought in future work.

The crossover example models were fitted with SAS software, and the visual impairment example models were fitted with R. In SAS, all models can be fitted with PROC NLMIXED, with the LLN exact method requiring Newton's method and Gaussian Quadrature coding within the call. In R, LLN and LN models were fitted utilizing R package `lnMLE` (Heagerty (1999), Comstock & Heagerty (1999)). The other four models of the visual impairment example were fitted with a modified `gnlmix()` R function of the `repeated` package (Lindsey (2001), see R code file). We note generally that: (1) it required substantially more quadrature points to obtain stable variance estimates than are currently used as defaults in existing packages (we used 100 points in the cross-over example); (2) alternative integration methods are available, such as MCMC, Romberg, etc. but are more computationally demanding. We used glnmix()'s exact Romberg integration for the visual impairment example and experienced exceedingly lengthy computational times; and (3) the LLN model requires two levels of integration, for each integral in constructing the marginal likelihood an additional integral must be performed for constructing the delta function; making it the most difficult to estimate.

Furthermore, the study design itself may have bearing on the appropriateness of conditional versus marginal estimates. In the cross-over study, direct observations of outcomes under both treatments are available for estimating the subject-specific parameters, since the predictor of interest (treatment) varies within a subject; this makes conditional interpretations more compelling for the cross-over trial example and MMMs straightforwardly provide the complementary marginal interpretations. In the visual impairment example however, only marginal information across groups is available on race differences. The data contain no directly observed information for the conditional, subject-specific question of how outcomes would differ if an individual's race changed, given the predictor does not vary within a subject. In this case, mixed-models produce conditional estimates that can be thought of as causal extrapolations. Though one may be interested in the counterfactual world describing expected outcomes which would have occurred if a particular person had been of an alternate ethnicity, no directly observed data are available for support. Estimability of the subject-specific effects here comes via the assumptions placed on the random effects and their distribution. If the often unverifiable random effects assumptions are incorrect, the conditional estimates can be far from the truth, as in any extrapolation. MMMs allow the use of mixed model architectures for describing associations within clusters while providing appropriate marginal estimates for predictors that do not vary within clusters, avoiding this type of causal extrapolation.

There are a variety of opportunities for further research. Given that MMMs are likelihood based, Bayesian castings could be investigated by adding prior distributions. Identifying additional conditional structures that produce commonly used marginal models will extend the availability of the marginalization approach, such as copulas. As further computational

advances are made in mixed model estimation, the connections shown here allow them to translate to advances in marginal model estimation as well. Performing sensitivity analyses towards latent variable assumptions will be furthered by having a wide range of easily implemented random effect distributions, as set forth by Lee & Thompson (2007) and Nolan (2010). Thus, investigations of alternate latent variable constructions via distributional transformations (as used in the LLB example) will be beneficial, as such transformations allow the random intercept to be on the same scale as the conditional coefficients, an issue raised in Lee & Nelder (2004).

Acknowledgements

References

Comstock, B & Heagerty, P (1999), 'Introduction to lnMLE: An R Package for Marginally Specified Logistic-Normal Models for Longitudinal Binary Data,' .

Heagerty, P (1999), 'Marginally specified logistic-normal models for longitudinal binary data,' *Biometrics*, **55**(3), pp. 688–698.

Heagerty, P & Zeger, S (2000), 'Marginalized multilevel models and likelihood inference,' *Statistical Science*, **15**(1), pp. 1–19.

Lee, K & Thompson, S (2007), 'Flexible parametric models for random-effects distributions,' *Statistics in medicine*, **27**(3), pp. 418–434.

Lee, Y & Nelder, J (1996), 'Hierarchical generalized linear models,' *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**(4), pp. 619–678.

Lee, Y & Nelder, J (2004), 'Conditional and marginal models: Another view,' *Statistical Science*, pp. 219–228.

Lindsey, J (2001), *repeated: Non-normal Repeated Measurements Models*, r package version 1.0.

Molenberghs, G & Verbeke, G (2005), *Models for discrete longitudinal data*, Springer Verlag.

Nolan, JP (2010), *Stable Distributions - Models for Heavy Tailed Data*, Birkhäuser, Boston, in progress, Chapter 1 online at academic2.american.edu/~jpnolan.

Robins, J, Rotnitzky, A & Zhao, L (1995), 'Analysis of semiparametric regression models for repeated outcomes in the presence of missing data,' *Journal of the American Statistical Association*, **90**(429), pp. 106–121.

Rubin, D (1976), 'Inference and missing data,' *Biometrika*, **63**(3), p. 581.

Scharfstein, D, Rotnitzky, A & Robins, J (1999), 'Adjusting for nonignorable drop-out using semiparametric nonresponse models (c/r: P1121–1146),' *Journal of the American Statistical Association*, **94**(448), pp. 1096–1120.

....

4   Stat **2012**, 00 1–4
Prepared using staauth.cls